A comprehensive comparison of students and non-students in classic experimental games

Citation for published version:
Belot, M, Duch, R & Miller, L 2015, 'A comprehensive comparison of students and non-students in classic experimental games', *Journal of Economic Behavior & Organization*, vol. 113, pp. 26–33.
https://doi.org/10.1016/j.jebo.2015.02.007

Digital Object Identifier (DOI):
10.1016/j.jebo.2015.02.007

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Journal of Economic Behavior & Organization

A comprehensive comparison of students and non-students in classic experimental games

Michele Belot* Raymond Duch† Luis Miller‡

January 16, 2015

Abstract

This study exploits the opening of the experimental lab in Oxford to compare the behavior of students and non-students in a number of classic experimental games, some of which involve other-regarding preferences (Trust Game, Dictator Game, and Public Goods Game) and others which have game forms that are strategically

*University of Edinburgh, School of Economics, Room 3.11, 31 Buccleuch Place, Edinburgh, UK. **Corresponding author**. Michele.Belot@ed.ac.uk

†Nuffield College, New Road, University of Oxford, Oxford, OX1 1NF, UK. raymond.duch@nuffield.ox.ac.uk

‡University of the Basque Country (UPV/EHU), School of Economics, Av. Lehendakari Aguirre 83, 48015 Bilbao, Spain. luismiguel.miller@ehu.es

challenging (Beauty-contest and Second-price Auction). We find that students are more likely to behave as selfish and rational agents than non-students. Our findings suggest that students are different than non-students with respect to their social preferences and their ability to reason strategically. Experiments using students are likely to overestimate the extent of selfish and rational behavior in the general population.

Keywords: laboratory experiments, subject pools, convenience samples

JEL classification: C72 C91

1 Introduction

Most subject pools in experiments conducted in social sciences are drawn from undergraduate student populations (Morton and Williams 2009; Peterson 2001). Based on a survey of 60 laboratory experimental economics papers published in major experimental economic journals, Danielson and Holm (2007) found that only four did not use students as subjects. Students are usually perceived as a "convenience sample" and preferred to non students for methodological reasons. We expect them to understand the instructions better; students are thought to be more comfortable with abstract reasoning; and they are also considered to be a more homogenous population. These advantages allow experimental researchers to obtain insights at relatively low costs and with small sample sizes. But, an important question is whether students differ in fundamental determinants of behavior in economic games. For example, do they differ in their ability to reason strategically, in their attitudes towards risk or in their pro-sociality? And does this matter for their behavior in experimental games?

We invited 63 students and 65 non-students to participate in the first experiment conducted in the Nuffield Centre for Experimental Social Sciences. The idea here is simply to compare two different samples that are available to experimentalists, neither of which can be claimed to be representative of the whole population. All participants played the same five classic experimental games; some of which involve other-regarding preferences (Trust Game, Dictator Game, and Public Goods Game); and others which have game forms that are strategically challenging (Beauty-contest and Second-price Auction). We identify the difference in behavior between the two subject pools in each of these games. To this, we will use two measures: One is the percentage of people playing the Nash equilibrium strategy based on monetary payoffs, assuming selfishness and rationality and common knowledge of rationality. The second measure is the average decision in each of the five games (e.g. average monetary transfer, average guess, etc.).

We find that the behaviour of the two subject pools can differ significantly although these differences vary across games. In games that involve other-regarding preferences, students are two to three times more likely to choose the equilibrium strategy than non-students. In contrast, the differences in behavior are smaller and not always significant in games that do not involve other-regarding preferences. We find no evidence that these differences are due to differences in levels of comprehension of the games. The differences are most dramatic for the simplest of the games (the Dictator Game), but also subsist when subjects have time to learn. The Public Goods Game was repeated 10 times and the differences are as pronounced at the beginning as at the end of the experiment. Finally, we find that behavior is correlated with individual covariates such as age, and cognitive ability in games involving cognitive reasoning.

The conclusion we draw from our study is that there are large differences in behavior across sub-groups of the population, but that these differences are more pronounced in games involving other-regarding preferences than in other games.

The rest of the paper is structured as follows. Section 2 presents the related literature, Section 3 the experimental design, Section 4 the results and Section 5 concludes.

2 Related literature

Despite the large reliance on students in laboratory work in economics, there is only a limited number of methodological studies comparing students to non-students (Carpenter et al. 2005; Carpenter et al. 2008; Anderson et al. 2013; Falk et al. 2013). Other social sciences (e.g. psychology, marketing) with a long tradition of experimental work have identified important behavioral differences between students and non-students, for example in personality tests, perceptions and attitudes, etc. A comprehensive meta-analysis conducted by Peterson (2001) suggests that student samples tend to be much more homogeneous than non-student samples and that the treatment effect sizes differ significantly across these two types of samples. However, we know relatively little about differences in behavior in economic games.

There are a few studies comparing pro-social behavior between students and other groups of the population. Carpenter et al. (2005) compares students and workers in both a laboratory and a field setting and finds that workers were more generous in the Dictator Game while students made more generous offers in the Ultimatum Game. A comparison of contributions of students and community members in a charitable donation version of the Dictator Game found that students were much less generous (Carpenter et al., 2008). More recently, Anderson et al. (2013) find that self-selected students appear considerably less pro-social than self-selected adults in a prisoner's dilemma game. Cappelen et al. (2015) show that a nationally representative adult sample behave more pro-socially than a student sample. Falk et al. (2013) find that non-students make significantly more generous repayments in a trust game. These authors conclude that measures of pro-social behaviour obtained from student convenience samples will be significantly lower than those from samples that are more representative of the broader population.

Next to these, there are a couple of studies comparing behavior in games that do not involve other-regarding preferences. A classic study is the comparison of newspaper and lab Beauty-contest experiments (Bosch-Domenech et al., 2002) which concludes that the lab and field generate comparable results. Depositario et al. (2009) find no difference in the bidding behavior of students and non-students in a uniform price auction.

Hence there is some evidence that the behavior of student and non-student subject pools differ. But the evidence remains very limited and typically focuses on one or two games. We have implemented a set of experiments designed to assess whether, and how, these subject pool differences vary across types of games.

3 Experimental design and sample

3.1 Description of samples and experimental sessions

The experiment was carried out in the Nuffield CESS lab in Oxford. It was the first experiment conducted in the lab. Hence, the subjects had not yet been exposed to experiments carried out in the lab. Students were recruited mainly among the undergraduate and graduate students at the two universities located in Oxford, while non students were recruited in various ways, such as e-mailing non-academic staff working at the university, contacting local shops, placing advertisements in local newspapers and local pubs.[1]

[1] The advertisement mentioned that we were looking for people from all walks of life to participate in decision-making experiments and surveys in social sciences. We also mentioned that participants should expect to earn between £10 and £15 per hour. The typical gross pay of students working for Oxford university is £12, and the average salary of an administrator in the UK in 2008 was £16,994, which corresponds to an hourly rate of £8.5 (Source: http://www.mysalary.co.uk/). Because the participants are informed about their expected payment, we would not expect systematic differences in the opportunity cost of participating. These two populations are both local and both prepared to participate in

We conducted six sessions: Two with students, two with non-students and two with a mixed population (a total of 128 subjects).[2] All sessions were conducted at 5.30 or 6 pm. Each session lasted for about one hour and a half.[3] Upon arrival, subjects were assigned randomly to a desk in the lab. We read out the instructions aloud at the beginning of each session.[4] The participants earned £10.7 on average.

Subjects were told the experiment had two parts. The first part involved 6 different situations (described in the next section), which we asked subjects to treat independently of each other. Earnings for this part were determined by selecting one of the six situations randomly (subjects were informed of the payment scheme at the very beginning). In the second part, participants were asked to do a short 12-minute version of an IQ test, consisting of 4 components: Numerical computation (calculation), numerical reasoning (e.g. logical series), abstract reasoning (figures and series) and verbal ability (e.g. analogies). Subjects had a limited amount of time for each component (2 or 3 minutes), and they received £0.20 for each correct answer.[5] Our measure of cognitive ability is a compromise

---

experiments. Thus, we are comparing two pools that could realistically be used in experiments conducted at Oxford.

[2] We found no evidence of differential play according to the session type. The analysis presented here pools the results across sessions.

[3] Sessions, especially those consisting mostly of non-students, lasted longer than expected and average payoffs were then lower than announced.

[4] The instructions can be found in the online supplementary material.

[5] Because of time constraints, we could not conduct a full IQ test. Instead, we selected 26 questions in total from an on-line psychometric test battery: www.psychometric-success.com. We sampled the

between a more complete measure (that would require a much longer time period) and more partial measures (measuring only certain dimensions of cognitive skills).

We also elicit risk preferences in the first part of the experiment (situation 5). Since we expect non-students to be quite heterogeneous in their mathematical ability, we chose a simple strategy method to elicit risk preferences (Dave et al., 2010). Participants are presented with eight successive choices between a safe amount (£5) and a lottery. The lottery gave them a 50% chance of earning a positive amount (which decreased by £1 in each case from £14 in the first case to £7 in the last case) and a 50% chance earning nothing. We told them that we would select one of the cases randomly to determine the payoff for this situation. Additionally, we have information on the subjects' age and gender.[6]

Table 1 presents summary statistics of the two samples. Students obviously are more homogenous in age than non-students. They also have a much higher average cognitive ability score, but there is a fair amount of variation both among students and non-students.

questions according to their levels of difficulty, including very easy, easy and more difficult questions. The level of difficulty is typically increasing in such tests, so we selected questions from the beginning, the middle and the end, to ensure we could capture well differences in cognitive ability across subjects. The test score ranges between 0 and 520, with 20 points attributed to each correct answer.

[6]We only have limited socio-economic information for non-students: 20 are public employees, 13 are private employees, 11 are self-employed, 4 unemployed, 8 are inactive and the remaining 9 have missing information. All students obviously share the same professional status. They come from a large variety of disciplines in social and natural sciences. Only 8 students were economics students.

In terms of risk preferences, measured by the number of lottery choices, we find that non-students tend to be less risk averse on average.

Summarizing, our subject pools do look different in important dimensions. Students are younger and smarter. These differences could be an advantage, insofar as they may help students better understand the instructions and the games played. But the differences could also compromise the generalizability of the results if they directly affect their ability to reason strategically or are correlated with risk or other-regarding preferences (Anderson et al., 2013).

Table 1. Main characteristics of the two samples

	Students			Non-students		
	Mean	Std. Dev.	Min-Max	Mean	Std. Dev.	Min-Max
Age (years)	23	4.7	18-35	34	11	18-70
Cognitive ability (0-520)	415	71.8	240-520	350	91	120-520
% female	54			63		
Number of lottery choices (0-8)	2.6	2.5	0-8	3.6	2.9	0-8

3.2 Description of the games

Each situation in the experiment (except for situation 5 that measured risk preferences) corresponded to a game. We read the instructions for the first situation; subjects were instructed to take a decision after we read the instructions. We then moved on to the second situation and so on. The situations were presented in the same order in each session. Subjects received no feedback about their payoffs in each of these situations until all six situations were completed (except in the last situation, which featured a repeated interaction).

The games played (in chronological order) were the following:

1. **Trust game:** We conducted a two-person binary version of the trust game (Berg et al 1995). Subjects are randomly paired and then we randomly assign one member of the pair to the role of 'sender' and the other to the role of 'receiver'. The sender received £10 and had to choose whether to transfer them to the receiver or to keep them. The money transferred was then tripled by the experimenter. The receiver then made a decision using the straegy method. Accordingly, the receiver was asked to decide whether, in the event that the sender sent the £10, she would keep the resulting £30 or split them evenly (£15- £15) between herself and the sender.

2. **Guessing game:** In each session all participants took part in the same guessing game. They had to guess the closest number to two-thirds of the average guess. The winner won £20, the rest nothing. In the event of ties, one of the winners was selected randomly.

3. **Dictator game:** We conducted a single-blind two-person dictator game. Subjects are randomly paired and then we randomly assign one member of the pair to the role of 'sender' and the other to the role of 'receiver'. The sender was told that she

had been allocated £10 and was asked to indicate the amount she wished to transfer to the receiver (the amount could be any integer number between 0 and £10).

4. **Second price sealed bid auction:** We conducted a second-price auction with all the subjects in one session bidding for a single unit of a commodity under a sealed-bid procedure. The highest bidder paid the price of the second highest bidder and earned profit equal to her valuation less the price paid. Other bidders earned zero profit. In the event of ties, one of the winners was selected randomly. Private valuations were assigned randomly with one-fourth of the subjects bidding for £4 and a similar proportion bidding for £6, £8, and £10.

5. **Repeated public good game:** Our public good game uses a standard Voluntary Contribution Mechanism (VCM) whereby the amount of public good produced is determined by the total contribution in the group (Isaac and Walker, 1988). The game has a unique equilibrium of full free-riding (dominant strategy in the one-shot game, unique Nash in the finitely repeated game). Our experimental environment mirrors those of previous experiments. Four subjects are endowed with 20 tokens. A subject can either keep these tokens for herself or invest them into a so-called 'project' by contributing g_i. The payoff function was the following:

$$\pi_i = 20 - g_i + 0.4 \sum_{j=1}^{4} g_j$$

The size of the project is just given by the sum of all contributions g_j to it. The marginal payoff of a contribution to the public good is 0.4 tokens.

The game was repeated 10 times with an anonymous partner matching design.

These five games vary in the degree to which they involve other-regarding preferences (which are likely to play a much larger role in the dictator game, trust game and repeated public good game) and in how easy they are to understand. We used the following method to measure the difficulty that subjects had understanding each of the games. Immediately after the instructions were read out for each game, we asked subjects to describe an example of a possible outcome. The goal was to obtain a measure of understanding without priming them in any particular way. For each game, we noted whether the example provided was correct or not. The ranking of the degrees of understanding of the games was similar for the student and non-student samples: The game that is best understood is the dictator game, this is followed by the trust game, the second price auction ranks third, the guessing game ranks fourth and the repeated public good game was least understood.

4 Experimental results

Table 2 compares the behavior of students (S) and non-students (NS). We present two measures: One is the percentage of people playing the Nash equilibrium strategy based on monetary payoffs -the homo-economicus equilibrium strategy, assuming selfishness and rationality and common knowledge of rationality (columns entitled "EB-S" relate to students and "EB-NS" relate to non-students). The second measure is the average decision in each of the five games (columns entitled "D-S" relate to students and "D-NS" relate to non-students)

For the Guessing Game, the Nash equilibrium is that subjects should report 0. But this is not very informative – only one subject does so in our sample. Alternatively, it is common for this game (Nagel, 1995) to define strategies according to depths of levels of reasoning. Level 0 subjects are non-sophisticated subjects and simply pick a number at random. Level 1 subjects best respond to Level 0 subjects whose random picks will average around 50, thus report a guess equal to 2/3 of 50. Level 2 subjects best respond to Level 1 subjects, thus report a guess equal to $(2/3)^2$ of 50. More generally, the level of depth of reasoning, k, commands a best response equal to $50 * (2/3)^k$. We will consider the probability of reasoning of level 1 or more.[7] The Nash equilibrium strategy for the other games are the following: Dictator Game – donate zero; Trust Game – send zero and send back zero; Public Goods Game – contribute zero. For the second-price sealed bid auction, the equilibrium strategy corresponds to bidding the subject's private value (Kagel and Levin, 1993).

In all games, with the exception of the auction game, we find that students are significantly more likely to behave as selfish and rational individuals.

Starting with the dictator game, we find that fifty-seven percent of the students donated

[7] The choice of level 1 is to some extent arbitrary, but we obtain very similar results if we compare the percentages of level 2 or level 3 play.

Table 2. Aggregate behavior in the five games

Game	Decision	EB-NS	EB-S	p-value	D-NS	D-S	p-value
Dictator	£0 - £10	17	57	p<0.001	3.5	1.6	p<0.001
Trust - sender	£0 - £10	18	65	p<0.001	8.2	3.5	p<0.001
Trust - receiver	£0 - £15	13	44	p=.005	13.1	8.4	p=0.005
Public Good r1	20 tokens MRC=0.4	9	24	p=0.026	11.1	9.9	p=0.323
Public Good r10	20 tokens MRC=0.4	29	52	p=0.008	5.4	4.8	p=0.151
Guessing	£20 if closest to 2/3 of average (0-100)	32	56	p=0.008	45.9	38.3	p=0.051
Auction	Private vale: £4	29	27	p=0.908	4.8	3.7	p=0.205
Auction	Private vale: £6	46	31	p=0.420	5.1	4.9	p=0.730
Auction	Private vale: £8	13	62	p=0.005	5.6	7.7	p=0.008
Auction	Private vale: £10	18	36	p=0.175	7.3	7.7	p=0.311

Note: p-values obtained from two-sample test of proportions and two-sample ranksum test.

EB-(N)S: % equilibrium behavior of non(students). D-(N)S: actual decision of (non)students.

nothing. In contrast, only seventeen percent of our non-students donated nothing which is less than a third of the student subject pool result. The average donation for our student sample was 16 percent; it was 35 percent for our non-student sample. Both of these differences are significant. The 2-1 non-student-student ratio in average donations is roughly the difference in donations that Carpenter et al. (2005) find between Middlebury students and Kansas City workers. Dictator game results for our student sample are also in line with the previous literature. Dictator games in which the recipients are anonymous result in average donations between 10 and 15 percent (Hoffman et al., 1994; Hoffman et al., 1996; Eckel and Grossman, 1996). With regard to the non-student sample results, Duch and Palmer (2004) find a similar reluctance of general population samples to make no donation – in their representative sample from Benin only eight percent made this choice.

Turning to the binary trust game, we find that only 35 percent of our student trustors trust in comparison to 82 percent of non student trustors. In their base-line treatment Berg et al. (1995) find that 90 percent of first movers invest; in a similar trust experiment Ortmann et al. (2000) and Eckel and Wilson (2004) find that 80 percent of first movers invest. In a binary trust experiment very similar to ours, Casari et al. (2007) report that 73 percent of their student subjects (designated as trustors) trusted by investing. Our non-student participants behavior resemble these results: Thus, it could be that having no experience as subjects in lab experiments (which was the case here) makes students particularly less trusting.

We also find significant differences in the rates of reciprocation: 56% of the student population and 87% of the non-student population reciprocated. Rates of reciprocation in binary trust games typically are in the range of 50-60 percent. Casari et al. (2007), for example, report a rate of reciprocation of 60 percent for their student subjects. Thus the student results are roughly consistent with reciprocation results in other studies while our non-student subjects here are clearly much more generous in their reciprocation.

Our third game involving other-regarding preferences is the public good game – this is the one game that involved repeated play (10 rounds). As table 3 indicates, students are considerably more likely to free-ride: in the first round of the public goods game, 24 percent of the students contributed nothing to the public good while only 9 percent of non-students were strictly non-cooperative. Gachter et al. (2004) conduct a similar, although one-shot, public good game with students and non-students. Their results suggest, consistent with ours, that non-students are more cooperative than students: non-students contributed an average of 10.2 tokens (out of 20) while students contributed 8.8. The average contribution of our non-students in the first round was 11.1 (out of 20) compared to 9.9 for students; in the last round the average contributions, respectively, were 5.4 and 4.8. On balance the patterns in both studies are similar and provide evidence that students are less cooperative than non-student subject pools.

Our two other games – the Auction and Beauty-contest games – involve no (or little) role for other-regarding preferences, since there is only one person winning the game and the others earn nothing. In these games students are more likely to make the equilibrium choice; but the magnitude of these differences is much smaller than it was for games involving other-regarding preferences. Table 2 suggests that for the Beauty-contest the ratio of student equilibrium choices to non-students equilibrium choices is roughly 2 to 1 compared to the 3-1 ratio for the other regarding games. And for the second-price sealed bid auction game the ratio is even smaller (1.5 to 1).

Turning to the beauty contest game, we again find significant differences between students and non students. The equilibrium strategy here is 0, yet typically the percentage of subjects choosing 0 in experiments is less than 10 percent (Camerer 1997; Nagel 1995) – only one non-student subject in our experiment selected 0. Thirty percent of our non-student subjects made choices consistent with at least level 1 iterative reasoning compared to 56 percent for the student subjects. The average guess of our student subjects is 38 which is consistent with similar guessing games results (in which the average is multiplied by two-thirds). Nagel (1995) finds that the average guess is 35. Camerer (1997) conducted multiple Beauty-contest experiments (with a multiple of .70 applied to the mean) and found the average guesses, for mostly student subject pools, ranged between 31 and 40. Our non-student subject pool exhibited considerably less iterative rationality with an average guess of 46, which seems indistinguishible from random behavior (an average of 50). On balance, these results are consistent with students exhibiting higher levels of iterative reasoning although the differences are less stark than with games involving other-regarding preferences.

In the second-price sealed bid auction, individuals should bid their private value (Kagel and Levin, 1993). Thirty-seven percent of our student subjects do bid their private value. Again this is consistent with other experimental results; Kagel and Levin (1993) find that 37 percent of students bid their private value in a second-price sealed bid auction. But our non-students are somewhat less likely to behave rationally and bid their private value – 26 percent of non-students behave in this fashion. But these differences are not statistically different.

Summarizing, the evidence so far suggests that students behave differently both in strategic and non-strategic games, but the differences are smaller in games that exhibit a more complex strategy form (repeated public good game and second price sealed bid auction).

Since the two populations are quite different from each other in many dimensions, it is perhaps not so surprising that they have quite different socio-economic profiles. Oxford students are younger, have better cognitive skills and are more likely to come from more privileged socio-economic backgrounds. Our goal here is not to identify the demographics or socio-economic characteristics that explain these differences. Our sample is too small to study this in detail. We can, however, see how robust the subject-pool effect is to the inclusions of various demographic variables. Specifically, we will control for measures of cognitive ability, age and gender. There is a body of empirical literature suggesting that these characteristics are correlated with behavior in a number of economic games.[8] Again,

[8]Burks et al. (2009) find that higher cognitive ability increases the probability of playing the non-cooperative strategy in a Prisoner's dilemma game. Sutter and Kocher (2007) find that trust increases almost linearly with age from early childhood to early adulthood, but stays rather constant between

to ease the comparison across games, we focus on the differences in equilibrium play.[9]

In Table 3, we estimate a series of probit regressions where the dependent variable is the equilibrium play dummy variable. Each cell in the table presents the estimate of the "student dummy" coefficient, capturing the difference in equilibrium play between students and non-students. Column (1) corresponds to estimates with no other covariates. Columns (2) to (4) include successively our measure of cognitive ability, age and gender as additional regressors. Overall we find that the subject-pool effect is robust to the inclusion of these additional variables for the games involving other-regarding preferences (rows (1) to (4), columns (1) to (4)). The estimated effect is not significant anymore in the guessing game once we control for age, while it becomes significant in the auction game once controlling for age as well (column (3)). One possibility is that age captures a form of ability, not captured by our measure of cognitive ability. This hypothesis seems to be supported by the fact that we do have a strong negative correlation (-.48), significant at the 1% level, between cognitive ability and age. A possible explanation is that older people may be less able to reason strategically.[10] Finally, the estimated effects are almost

different adult age groups. A large number of experimental studies in economics and psychology have documented gender differences in preferences (Croson and Gneezy, 2009).

[9]Similar results are obtained if we use the continous decision variables, results are available upon request.

[10]However, this explanation must be viewed with caution. Selection into experiments may have different consequences in different age groups and recruiting high cognitive ability young people might be much easier than recruiting high cognitive ability middle-aged people, who may have better jobs and less time and need to come to the lab.

identifical when controlling for gender (column (4)).

In addition to these individual characteristics, we also collected information on risk preferences. In principle, there is only one situation in which risk attitudes might play a role - the trust game. Students are on average more risk averse, but the difference is small and not statistically significant. Note that the measure is noisy and shows a high degree of inconsistency in responses to the risk items given the simplicity of the question: About a third of the subjects (in both pools) switch back and forth (at least once) between the 'safe' and the 'risky' alternative.[11] With this caveat in mind, we restrict our sample to those who gave consistent answers. We find that including the measure of risk aversion hardly changes the coefficient of the student dummy in the trust regression, and the measure of risk aversion does not have a significant effect on behavior. For the sake of brevity, we do not report the results in a table.[12] The estimated student dummy hardly changes when controlling for risk aversion. Thus, the subject pool differences we estimate are robust to differences in risk aversion measured in this standard and simple way.

For the most part the subject pool differences in behaviour that we estimate are robust to controling for key demographic variables. This is certainly the case with the other-regarding games – the student and non-student differences are unaffected by the inclusion

[11]Non-unique switching points is not uncommon in these types of experiments. Holt and Laury (2002) elicit risk preferences in a similar fashion and report 19.9 percent of the subjects exhibited non-unique switching behavior.

[12]The results are available upon request from the authors.

Table 3 - Demographics and behaviour - Probability of equilibrium decision

Probit estimates - Estimated student dummy coefficient

	(1)	(2)	(3)	(4)
(1)Dictator game	.40***	.41***	.31*	.31*
	(.11)	(.12)	(.16)	(.16)
(2) Trust game - sender	.46***	.50***	.55***	.55***
	(.11)	(.12)	(.14)	(.14)
(3) Trust game - receiver	.31***	.22*	.38**	.38***
	(.11)	(.12)	(.14)	(.14)
(4) Public good game	.15**	.10	.10	.10
(round 1)	(.06)	(.07)	(.08)	(.08)
(5) Guessing game	.23***	.19**	.06	.06
	(.09)	(.09)	(.11)	(.12)
(6) Auction	.12	.11	.21*	.21*
	(.08)	(.09)	(.10)	(.10)
Control for cognitive ability	No	Yes	Yes	Yes
Control for age	No	No	Yes	Yes
Control for gender	No	No	No	Yes

Note: Each cell presents the estimates of the student dummy corresponding to a separate regression. ***, ** and * denote significance levels at 1%, 5% and 10%.

of demographic controls in the estimation. For the games require higher strategic reasoning the results are robust to most demographic variables with the exception of age – it would seem that older subjects exhibit lower strategic reasoning abilities.

5 Discussion and conclusion

Our results suggest that in games that invoke other-regarding considerations, student and non-student subject pools will make choices that are very different - students are more likely to behave as a selfish and rational agent than is the case for non-students. The student subject pool was less kind and less trusting than the non-student subjects. Fairly consistently across experiments, non-students were more other regarding by a ratio of three-to-one.

Student and non-student subject pools can differ quite significantly in terms of characteristics that likely matter for the decisions they take in a typical economic experiment, such as cognitive skills and age for example. In our study, students are better able to reason strategically and they are less pro-social. Of course we are looking here at two atypical samples, and none of them can be claimed to be a representative sample of the population. What our study shows is that there are large differences in behaviour across sub-groups of the population, but that these differences are more pronounced in games involving other-regarding preferences than in other games.

Of course, often experimentalists are not directly interested in "levels", but are rather interested in the effects of a particular treatment on behaviour. If a population differs in fundamentals driving behavior such as preferences and the ability to reason strategically, then most economic models would predict that treatment effects - changes in behavior in response to changes in the incentive structure - will differ as well, at least in a quantitative sense. So, for example, the introduction of punishment in a Public Goods game may induce higher contributions from students than non-students. Our results show that important socio-demographic characteristics (being a student, cognitive ability, IQ, age, etc.) are correlated with important fundamentals driving behavior. This raises the question of the implications the use of convenience samples have for the conclusions drawn from an experiment where we expect demographic characteristics to condition the treatment effect.

In many cases though we are interested in qualitative treatment effects – simply does punishment induce more cooperation? Hence, Kessler and Vesterlund (2011) argue that for most laboratory studies it is only relevant whether the qualitative results of the study are externally valid. If this is the case then it is not clear that using a student subject pool may compromise the generalisability of the findings in a qualitative sense. The qualitative effect – for example of punishment in Public Goods games – would be exhibited by both student and non-student subjects.

While we cannot provide a general test of the external validity of treatment effects, our design allows us to provide some insights into qualitative differences. These qualitative effects are not directly related to treatment effects but they suggest that the qualitative effects of different experimental conditions are similar for both students and non-students. First, we can compare levels of equilibrium play across games of different levels of complexity (i.e. the Dictator Game is less cognitively demanding than the Public Goods game). Second, we can compare equilibrium play across repeated rounds with feedback in the public goods game. We see that for both subject pools, the levels of equilibrium play are lower in more complex games. Considering for example only the games involving other-regarding preferences, we find that equilibrium play is lower in the Public Goods game than in the Trust game or Dictator game, the latter being arguably less complex. Similarly, we also find that the levels of equilibrium play increase for both students and non-students in the repeated Public Goods game when comparing round 10 to round 1. These results provide suggestive evidence that qualitative differences may be less affected by subject-pool characteristics. Of course these are not proper "treatments" and are only limited evidence, but the results are encouraging for studies mainly interested in identifying qualitative treatment effects.

Acknowledgments

We thank Dirk Engelmann and an anonymous referee for helpful comments and suggestions. Luis Miller acknowledges the financial support received from the Spanish Ministry of Economy and Competitiveness (grant ECO2012-30626) and the Basque Government

(research group IT-783-13).

References

Anderson, J., Burks, S., Carpenter, J., GÃ¶tte, L., Maurer, K., Nosenzo, D., Potter, R., Rocha,K., Rustichini, A., 2013. Self-selection and variations in the laboratory measurement of other-regarding preferences across subject pools: evidence from one college student and two adult samples. Experimental Economics 16(2), 170-189.

Berg, J., Dickhaut, J., McCabe, K., 1995. Trust, Reciprocity, and Social-History. Games and Economic Behavior 10, 122-142.

Bosch-Domenech, A., Montalvo, J., Nagel, R., Satorra, A., 2002. One, Two, (Three), Infinity...: Newspaper and Lab-Beauty Experiments. American Economic Review 92, 1687-1701.

Burks, S., Carpenter, J. GÃ¶tte, L., Rustichini, A., 2009. Cognitive Skills Affect Economic Preferences, Social Awareness, and Job Attachment. Proceedings of the National Academy of Science 106(19), 7745-7750.

Camerer, C., 1997. Progress in Behavioral Game Theory. The Journal of Economic Perspectives 11, 167-188.

Cappelen, A., Nygaard, K., Sorensen, E., Tungodden, B., 2015. Efficiency, equality and reciprocity in social preferences: A comparison of students and a representative population. Scandinavian Journal of Economics. Forthcoming.

Carpenter, J., Burks, S., Verhoogen, E., 2005. Comparing Students to Workers: The Effects of Social Framing on Behavior in Distribution Games. Research in Experimental Economics 10, 261-290.

Carpenter, J., Connolly, C., Myers, C., 2008. Altruistic Behavior in a Representative Dictator Experiment. Experimental Economics 11, 282-298.

Casari, M., Ham, J., Kagel, J., 2007. Selection Bias, Demographic Effects, and Ability Effects in Common Value Auction Experiments. American Economic Review 97, 1278-1304.

Croson, R., Gneezy, U., 2009. Gender Di erences in Preferences. Journal of Economic Literature 47, 448-74.

Danielson, A. J., Holm H. J., 2007. Do You Trust Your Brethen? Eliciting Trust Attitudes and Trust Behavior in a Tanzanian Congregation. Journal of Economic Behavior and Organization 62, 255-271.

Dave, C., Eckel, C., Johnson, C., Rojas, C., 2010. Eliciting risk preferences: When is simple better? Journal of Risk and Uncertainty 41(3), 219-243.

Depositario, D., Nayga, R., Wu, X., Laude, T., 2009. Should Students be used as Subjects in Experimental Auctions. Economics Letters 102, 122-24.

Duch, R., Palmer, H., 2004. Its Not Whether You Win or Lose, but How You Play the Game: Self-Interest, Social Justice, and Mass Attitudes toward Market Transition. American Political Science Review 98, 437-452.

Eckel, C., Grossman, P., 1996. Altruism in Anonymous Dicatator Games. Games and Economic Behavior 16, 181-191.

Eckel, C., Wilson, R., 2004. Is Trust a Risk Decision? Journal of Economic Behavior and Organization 55, 447-466.

Falk, A., Meier, S., Zehnder, C., 2013. Do Lab Experiments Misrepresent Social Preferences? The case of self-selected student samples. Journal of the European Economic Association 11(4), 839-852.

Gachter, S., Herrmann, B., Thoni, C., 2004. Trust, Voluntary Cooperation, and Socio-Economic Background: Survey and Experimental Evidence. Journal of Economic Behavior and Organization 55, 505-531.

Hoffman, E., McCabe, K., Smith, V., 1994. Preferences, Property Rights, and Anonymity in Bargaining Games. Games and Economic Behavior 7, 346-380.

Hoffman, E., McCabe, K., Smith, V., 1996. Social Distance and Other-Regarding Behavior in Dicatator Games. American Economic Review 86, 653-60.

Holt, C., Laury, S., 2002. Risk Aversion and Incentive Effects. American Economic Review 92, 1644-1655.

Isaac, R. M., Walker, J., 1988. Group Size Effects in Public Goods Provision: The Voluntary Contributions Mechanism. Quarterly Journal of Economics 103, 179-99.

Kagel, J., Levin, D., 1993. Independent Private Value Auctions: Bidder Behavior in First-, Second- and Third-Price Value Auctions with Varying Numbers of Bidders. The Economic Journal 103, 868-79.

Kessler, J., Vesterlund, L., 2013. The External Validity of Laboratory Experiments: Qualitative rather than Quantitative Effects. In: Frechette, G., Schotter, A. (Eds.), The Methods of Modern Experimental Economics, Oxford University Press, Oxford.

Morton, R., Williams, K., 2009. From Nature to the Lab: Experimental Political Science and the Study of Causality, Cambridge University Press, New York.

Nagel, R., 1995. Unraveling in Guessing Games: An Experimental Study. American Economic Review 85, 1313-1326.

Ortmann, A., Fitzgerald, J., Boeing, C., 2000. Trust, Reciprocity, and Social History: A Re-examination. Experimental Economics 3, 81-100.

Peterson, R. A., 2001. On the Use of College Students in Social Science Research: Insights from a Second-Order Meta-analysis. Journal of Consumer Research 28, 450 461.

Sutter, M., Kocher, M. 2007. Trust and Trustworthiness across Different Age Groups. Games and Economic Behavior 59, 364-82.